# Simulation of Narrow Band Speech Signal using BPN Networks

G.Gandhimathi
Assistant Professor
School of Architecture Engineering and Technology
Periyar Maniammai University
Vallam, Thanjavur-613403, India

C.Narmadha
Assistant Professor
School of Architecture Engineering and Technology
Periyar Maniammai University
Vallam, Thanjavur-613403, India

C.Lakshmi
Assistant Professor
Department of ECE
School of Electrical & Electronics Engineering
SASTRA University

## ABSTRACT
This paper proposes to extend the band width of narrow band telephone speech signal by employing feed forward back propagation neural network. There are different types of faster training algorithm are available in the literature like Variable Learning Rate, Resilient Back propagation, Polak-Ribiére Conjugate Gradient , Conjugate Gradient with Powell/Beale Restarts , BFGS Quasi-Newton , One-Step Secant , Fletcher-Powell Conjugate Gradient Algorithms, Scaled Conjugate Gradient and Liebenberg-Marquardt. These algorithms are used to train the BPN networks using Neural network tool box. The correlation between the inputs of the neural network and the input-output correlation were calculated. The components were employed to reconstruct the speech signal and the results are analyzed.

## General Terms
Artificial band width expansion, Signal processing, Neural networks , Training algorithm .

## Keywords
AR Filter, Back propagation neural network , linear mapping method, code book method.

## 1. INTRODUCTION
The non availability of infinite channel bandwidth offered by any transmission media forces the researcher to go for sampling. This bandwidth limitation of the transmitted speech in current public telephone systems is due to the constraints of the old analogue telephone system to a frequency range of up to about 3.4 kHz. The compromise between the voice quality and the sampling rate stated the sampling rate at 8 kHz is suitable for most telephone systems. The voice signal, which occupies a band from 100 Hz to 8 kHz, is filtered to 4 kHz, to preserve the integrity of the signal sampled at 8 kHz. In this process, the frequency components from 4 to 8 kHz are excluded in the narrowband signal, and are known as lost or missing components. The components from 100 to 300 Hz and from 3400 to 4000 Hz are still present, although attenuated. The missing frequency components compromise the speech intelligibility.

The quality of the narrowband speech is acceptable for vowels, but is poor for consonants, mainly fricative consonants (Isl, /sh/, /ch/, xl, /th/, etc.). S. Chennoukh, *et al* [1] discuss the speech enhancement via frequency *b*andwidth extension using line spectral frequencies and there are several methods to estimate the lost components, e.g. linear mapping [2], neural networks and codebooks [3]. Furthermore Jax and Vary [4, 5] investigates the potential features and evaluate their suitability for the BWE application. The quality of each feature is quantified in terms of the statistical measures of mutual information and separability. It turns out that the best BWE results are obtained by using a large feature "super-vector" (i.e high mutual information) which is subsequently reduced in dimension by a linear discriminant analysis. Their solution also helps to reduce the computational complexity of the estimation of the wideband spectral envelope.

Miet *et al* [6] described a technique which splits the telephone-band speech signal into a spectral envelope and a short-term residual. Where the spectral envelope and the residual are extended separately and recombined to create an extended band signal. This system is evaluated by listening tests and distortion measurement. Qian and kabal presents [7] a novel approach which combines equalization and estimation to create a wideband signal, with reconstructed components in the 3400 Hz to 7000 Hz range. Equalization is used in the 3400-4000 Hz range. Its performance is better than statistical estimation procedures, because the mutual dependencies between the narrowband and high band parameters are not sufficiently large.

This paper organized as follows the section 2 explores the proposed methodology. Section 3 discuss in detail about the design of the proposed neural network and their performance. section 4 elaborates the results. Section 5 concludes a suitable training method.

## 2. THE PROPOSED METHOD
In this paper we treat the problem by processing the signal (4KHz) the narrowband signal up sampled by a factor of two (8KHz), into 16 channels or frequency bands. The first 7 bands corresponds the frequencies 0 up to 3500Hz. Those components can be found in the narrowband speech, and the other 9 channels are the ones missing in this signal. A feed forward neural network is used to map the band limited signal components into the missing components in a non-linear way. The signal is then reconstructed, as shown in Fig 1.

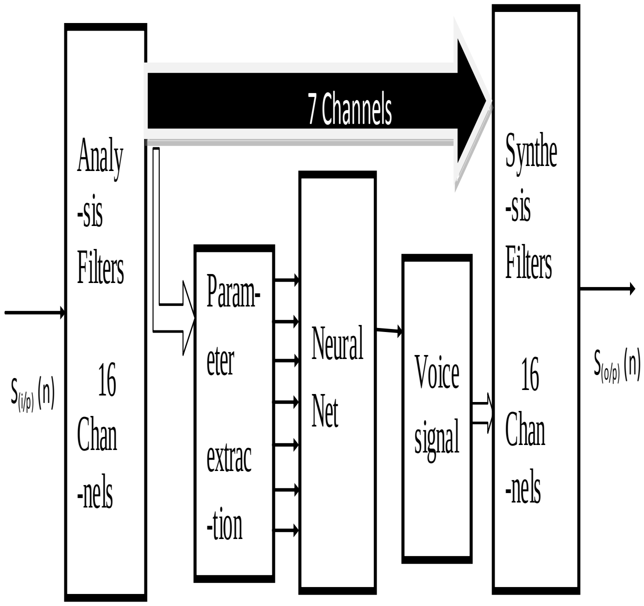


Figure 1.Frequency extension using neural Network

## 3. DESIGNING THE NEURAL NETWORK

### 3.1. Input and Output pairs

We used six phrases in English, with approximately three seconds each, sliced in 20 ms segments, to train the neural network. The input of the neural network was the narrowband speech coefficient of a first order AR filter for each of the seven first channels, and the output, were an estimation of the same parameter for the other nine channels. The architecture is shown in Fig2

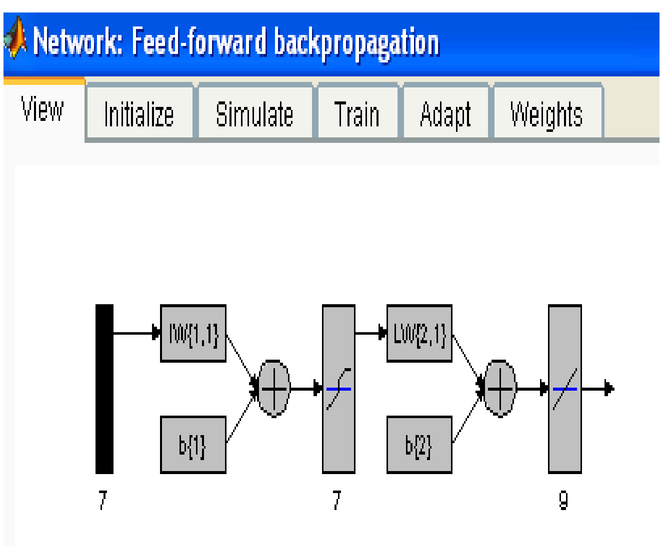


Fig 2 – Architecture of feed forward BPN Neural network

Table 1 - Correlations between inputs

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1 | 1.0000 | 0.0000 | 0.5731 | -0.0017 | 0.0755 | -0.4061 | -0.5915 |
| 2 | 0.0000 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | -0.0000 |
| 3 | 0.5731 | 0.0000 | 1.0000 | 0.0087 | 0.1446 | -0.4197 | -0.8908 |
| 4 | -0.0017 | 0.0000 | 0.0009 | 1.0000 | 0.0002 | 0.0010 | -0.0002 |
| 5 | 0.0755 | 0.0000 | 0.1446 | 0.0002 | 1.0000 | -0.0344 | -0.1228 |
| 6 | -0.4061 | 0.0000 | 0.4197 | 0.0097 | -0.0344 | 1.0000 | 0.6905 |
| 7 | -0.5915 | -0.0000 | -0.8908 | -0.0022 | -0.1228 | 0.6904 | 1.0000 |

Table 2 - Correlations between input and outputs

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1 | -0.5903 | -0.0000 | -0.8343 | -0.0000 | -0.1133 | 0.7268 | 0.9872 |
| 2 | -0.1235 | 0.0000 | -0.4946 | -0.0002 | -0.0515 | 0.8704 | 0.7587 |
| 3 | -0.5442 | 0.0000 | -0.4561 | 0.0008 | -0.0446 | 0.8257 | 0.5867 |
| 4 | 0.5587 | 0.0000 | 0.9540 | 0.0010 | 0.1494 | -0.4315 | -0.9195 |
| 5 | -0.5487 | -0.0000 | -0.8760 | -0.0004 | -0.1214 | 0.7506 | 0.9763 |
| 6 | 0.5405 | 0.0000 | 0.7235 | 0.0004 | 0.0868 | 0.8208 | 0.9466 |
| 7 | -0.7240 | -0.0000 | -0.8770 | 0.0003 | -0.1188 | 0.6940 | 0.9840 |
| 8 | 0.5970 | 0.0000 | 0.9320 | 0.0005 | 0.1340 | 0.6149 | 0.9912 |
| 9 | 0.0900 | 0.0000 | 0.4113 | 0.0023 | 0.0386 | 0.2490 | -0.1579 |

### 3.3. Training the neural network

The database consisted of 500 narrowband and wideband speech pairs of 20ms. The training, validation and test sets were composed by 300, 100 and 100 pairs, respectively. The training procedure was error back propagation supervised to avoid overtraining. At each 10 epochs the validation set was applied, and we considered that overtraining was occurring if the validation set error increased over 50 epochs.

We use different types of training algorithm and the corresponding training curves are shown in Figs 3 to 11. Audible tests were performed and confirmed this result.

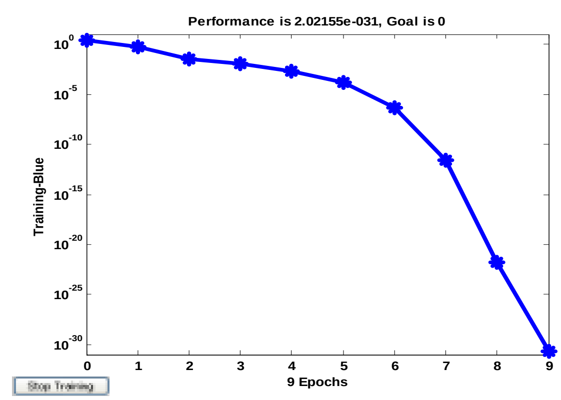


Fig 3-Training with Levenberg-Marquardt (trainlm) algorithm

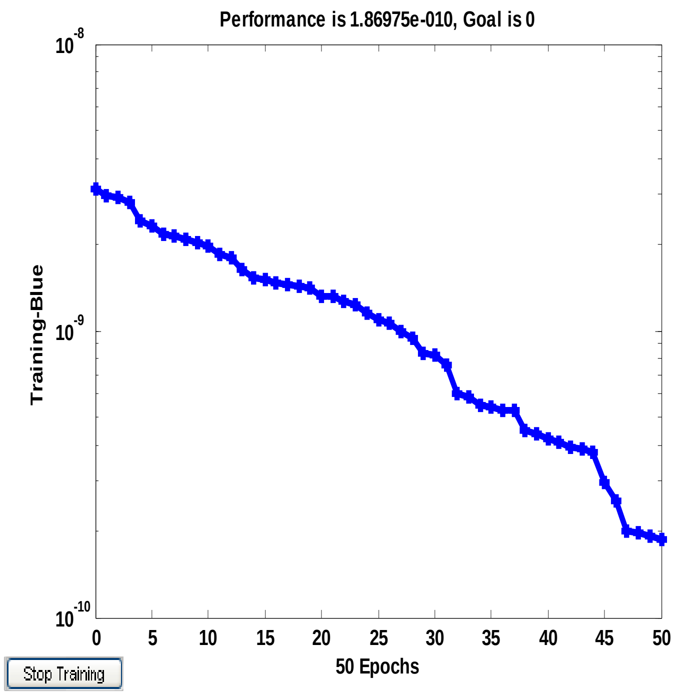


Fig 4- Training with Scaled Conjugate Gradient (trainscg) algorithm

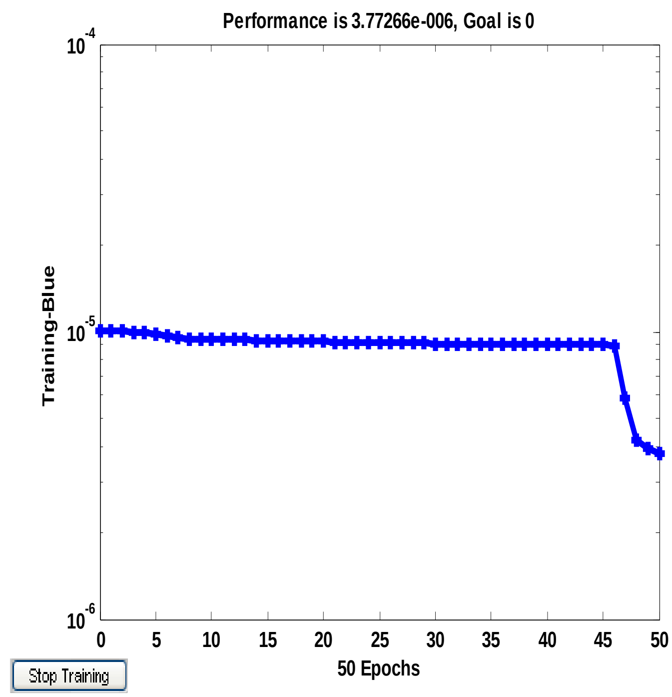


Fig 6- Training with One-Step Secant (trainoss) algorithm

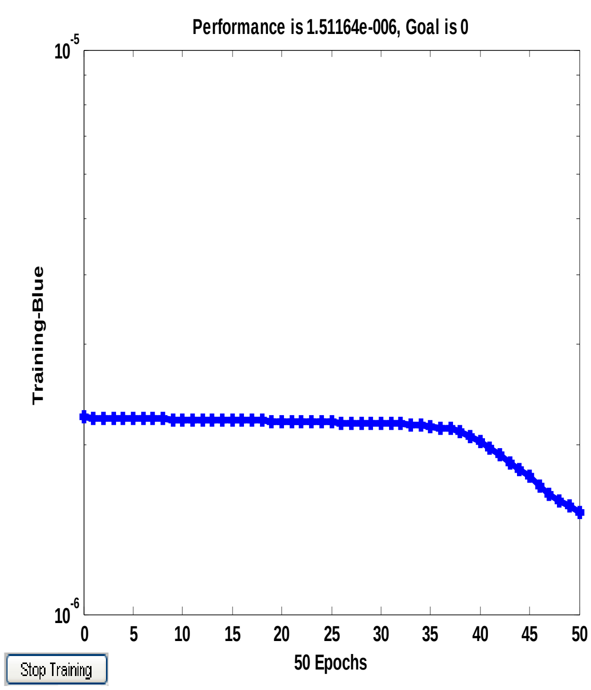


Fig 5- Training with Fletcher-Powell Conjugate Gradient (traincgf) algorithm

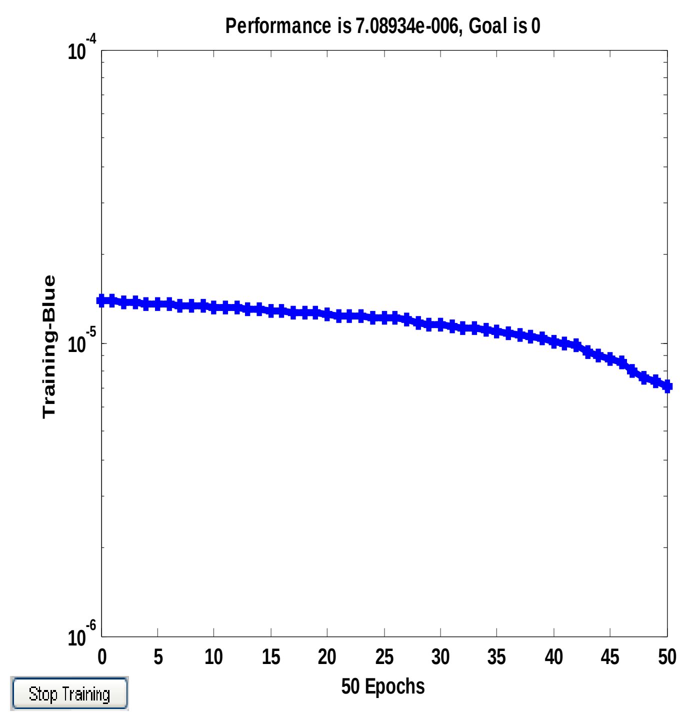


Fig 7- Training with BFGS Quasi-Newton (trainbfg) algorithm

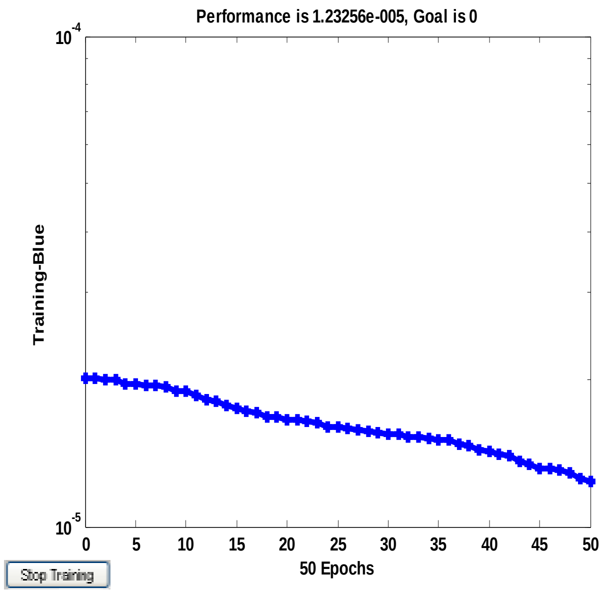


Fig 8- Training with Conjugate Gradient with Powell/Beale Restarts(traincgb) algorithm

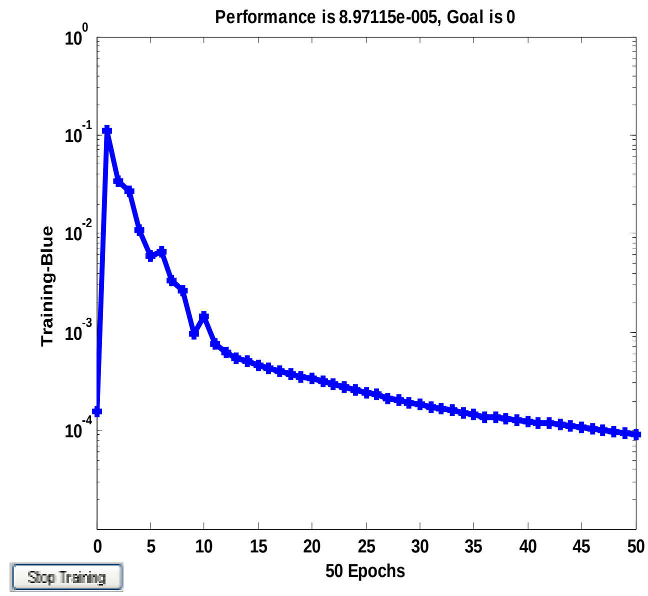


Fig 10- Training with Resilient Back propagation (trainrp) algorithm

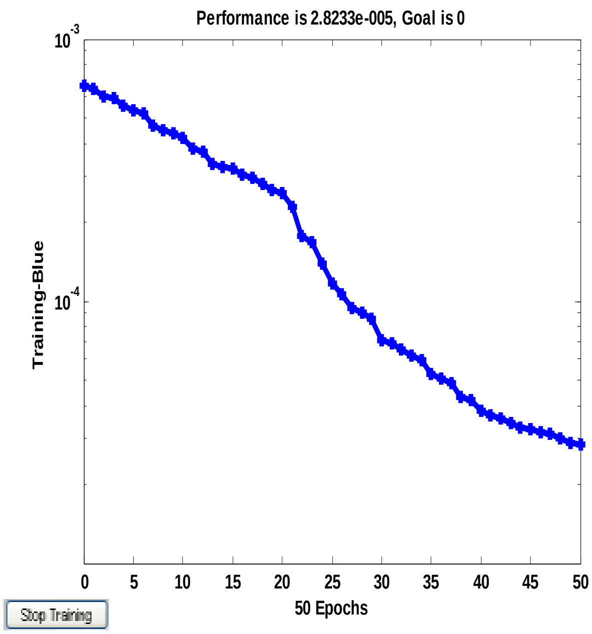


Fig 9- Training with Polak-Ribiére Conjugate Gradient (traincgp) algorithm

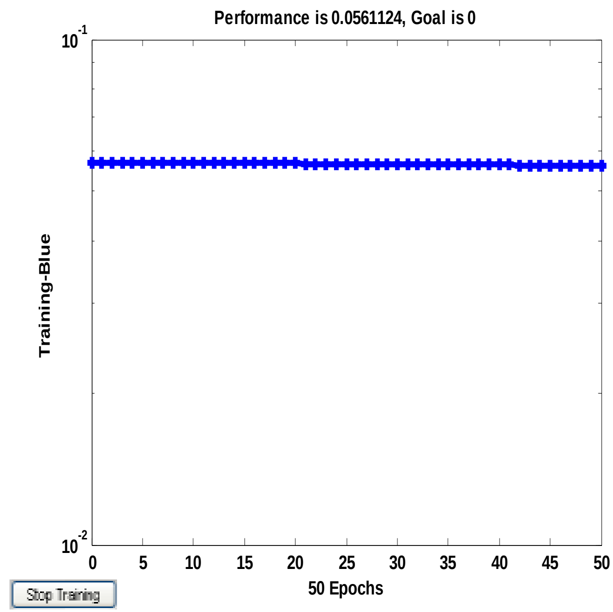


Fig 11- Training with Variable Learning Rate Backpropagation (traingdx) algorithm

## 4. RESULTS & DISCUSSION

The phrases used in the final test were: The fruit peel was cut in six slices, spoken by a female speaker. The narrow band speech signals are given to the BPN networks. we use the different types algorithm to train the network and its performance error is measured and displayed in Table 4 . The Fig 3 shows the training with Levenberg-Marquardt (trainlm) algorithm and reveils that it wolud take less error of 2.02155e-031within 9 iterations. Fig 4 shows the Training with Scaled Conjugate Gradient (trainscg) algorithm even after 30 iterations the error is above 1.86975e-010. From Fig 5 Training with Fletcher-Powell Conjugate Gradient (traincgf) algorithm the error is more but learn within 35 iterations, Fig 6- Training with One-Step Secant (trainoss) algorithm atleast 45 iterations , Fig 7,8,9 - Training with BFGS Quasi-Newton (trainbfg) algorithm, Training with Conjugate Gradient with Powell/Beale Restarts(traincgb) algorithm , Training with Polak-Ribiére Conjugate Gradient (traincgp) algorithm and Fig 10- Training with Resilient Back propagation (trainrp) algorithm needs more than 50 iterations to converge . Fig 11- Training with Variable Learning Rate Backpropagation (traingdx) algorithm needs more than 300 iterations and error is also more .

Table 4 Error performance for different algorithms

| **Acronym** | **Algorithm** | **Performance** |
|---|---|---|
| LM | Levenberg-Marquardt | 2.02155e-031 |
| SCG | Scaled Conjugate Gradient | 1.86975e-010 |
| CGF | Fletcher-Powell Conjugate radient | 1.40904e-006 |
| OSS | One-Step Secant | 3.77266e-006 |
| BFG | BFGS Quasi-Newton | 7.08934e-006 |
| CGB | Conjugate Gradient with Powell/Beale Restarts | 1.23256e-005 |
| CGP | Polak-Ribiére Conjugate Gradient | 2.8233e-005 |
| RP | Resilient Back propagation | 8.97115e-005 |
| GDX | Variable Learning Rate Back-propagation | 0.0561124 |

## 5. CONCLUSIONS

The Levenberg-Marquardt algorithm is the fastest training algorithm for networks of moderate size. It has memory reduction feature for use when the training set is large. This paper showed that a neural network (BPN with LM) can be used to conveniently solve the problem. We obtained good results, in terms of extension and quality of the reconstructed speech, and the method holds less computational complexity when compared to others the codebook method [3].

## 6. Acknowledgement

**The author wish to thank Prof. R.Amirtharajan Assistant Professor / ECE School of Electrical & Electronics Engineering, SASTRA University for his valuable guidance and support**

## 7. REFERENCES

[1] S. Chennoukh, A. Gerrits, G. Miet, and R. Sluitjer, "Speech enhancement via frequency *b*andwidth extension using line spectral frequencies," Pro. IEEE Int. Conf. On Acoustics, Speech, Signal Processing, vol. 1, pp. 665-668, May 2001.

[2] D.A. Heide, G.S. Kang, "Speech enhancement for band limited speech", Proceedings of the ICASSP, Vol. 1, Seattle, WA, USA, May 1998, pp. 393-396.

[3] B. Iser and G. Schimidt, "Neural Network Versus Codebooks in an Application for Bandwidth Extension of Speech Signal", in Proc. ROSPEECH'03, vol.1, pp. 565-568, September 2003.

[4] P. Jax and P. Vary, "An upper bound on the quality of artificial bandwidth extension of narrowband speech signals", Pro. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) vol. 1, pp 237-240, Orlando, FL, USA, May 2002.

[5] P. Jax and P. Vary, "Feature Selection for Improved Bandwidth Extension of Speech Signals", Pro. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) vol. 1, pp. 697-700, Montreal, Canada, May 2004.

[6] G. Miet, A. Gerrits, and J. C. Valiere, "Low-band extension of telephone-band speech," Pro. IEEE Int. Conf. On Acoustics, Speech, Signal Processing, vol. 3, pp. 1851-1854, Jun 2000.

[7] Y. Qian and P. Kabal, "Combining Equalization and Estimation for Bandwidth Extension of Narrowband Speech", Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (Montreal, QC), pp. 1-713-1-716, May 2004